Hybrid Cloud Solutions

Hybrid Cloud Solutions

NetApp
February 11, 2024

Table of Contents

Hybrid Cloud Solutions

The Hybrid Cloud Solutions page takes a solutions level approach to enabling customers on their cloud journey.

Top Links

- NetApp And PetaGene: Accelerating the Genomics Revolution
- Durability, Availability, Encryption and Snapshots on Cloud Volumes Service
- NetApp Cloud Volumes ONTAP And Enhanced Networking Support
- GDPR and Hybrid Cloud: A More Perfect Union

Find more information

- NetApp Cloud Central
- Get a first look at the new NetApp Cloud Volumes Service for AWS

Technology

Cloud Volumes Service

Blogs

NetApp Cloud Volumes, Not Your Mothers File Service

Authors: Chad Morgenstern, Karthik Chinnathambi

If you are thinking about your data in the clouds, consider that NetApp Cloud Volumes Service is about more than just NFS or SMB. With the Cloud Volumes Service, you can rest assured that your data is durable, encrypted, and highly available. But what about protection from user errors, wouldn't you love the ability to roll back time to undo choices you may have made and wish you hadn't? Yeah, we got you covered there too with NetApp Cloud Volume Service's Snapshot capability.

Durability, Availability, Encryption and Snapshots

Durability

Your data is protected not just against multiple drive failures, but against numerous type of disk errors that could otherwise affect not just your data durability but your data integrity as well. No more worries that your data is going to disappear.

Encryption

Regarding encryption, NetApp Cloud Volume service utilizes Storage Encryption to provide full-disk encryption without compromising storage application performance. This single-source solution can increase overall compliance with industry and government regulations without compromising your user experience.

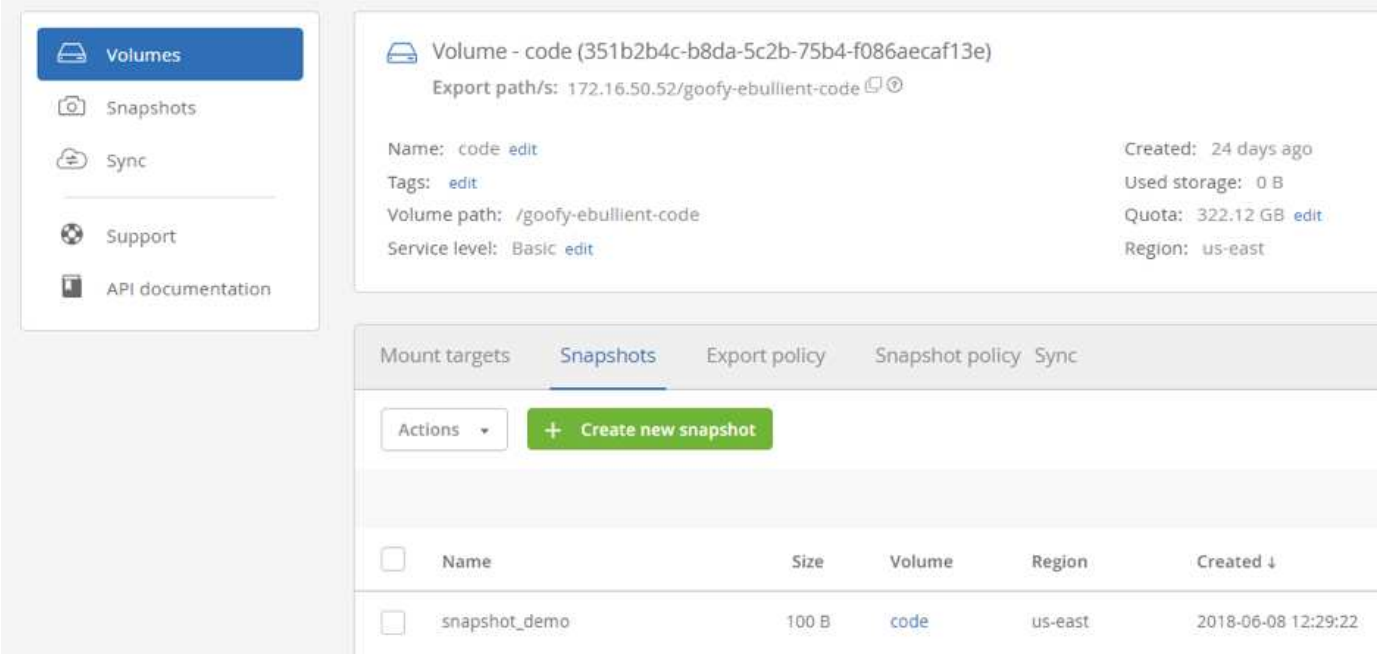
Availability

The Cloud Volumes Service is built upon best of industry, proven rock-solid enterprise class hardware and software all but ensuring uptime – availability. Each storage node has many paths to its own Solid State Drives protecting against path failure as well as access to a partner storage node SSDs as well. As each storage node has access to its partners disks, continued productivity in the event of controller failure is ensured. You are protected against network failures by redundant network ports and paths all the way up the stack across the cloud connections.

As NetApp Cloud Volume Service for AWS sits centrally in relation to each of the availability zones within an AWS region, the service is immune to availability zone outages. Access your data from any availability zone without the need to replicate content – the user experience from AZ to AZ is expected to be the same and if not, well, its outside of our control – clouds will be clouds after all.

The Undo Button

Snapshots act as logical backups and are as such point-in-time representations of your data. The creation of Snapshots may be scheduled or taken manually both within the Cloud Volumes Portal.



The Cloud Volume Service is evolving rapidly and as such I can tell you that API support for Snapshot creation is coming.

NetApp Cloud Volume Snapshots are fast, plentiful and non-disruptive. A NetApp Snapshot simply manipulates block pointers, creating a “frozen” read-only view of a volume that lets applications access older versions of files and directory hierarchies without special programming. Because the actual data blocks aren’t copied, Snapshot copies are extremely efficient both in the time needed to create them and in storage space. A Cloud Volume Snapshot copy takes only a few seconds to create—typically less than one second, regardless of the size of the volume or the level of activity within the environment.

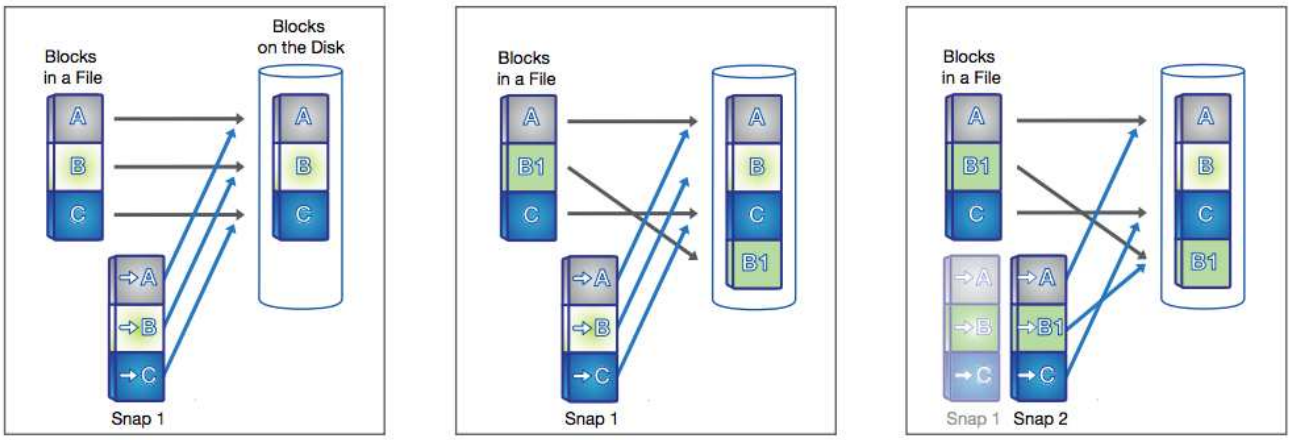
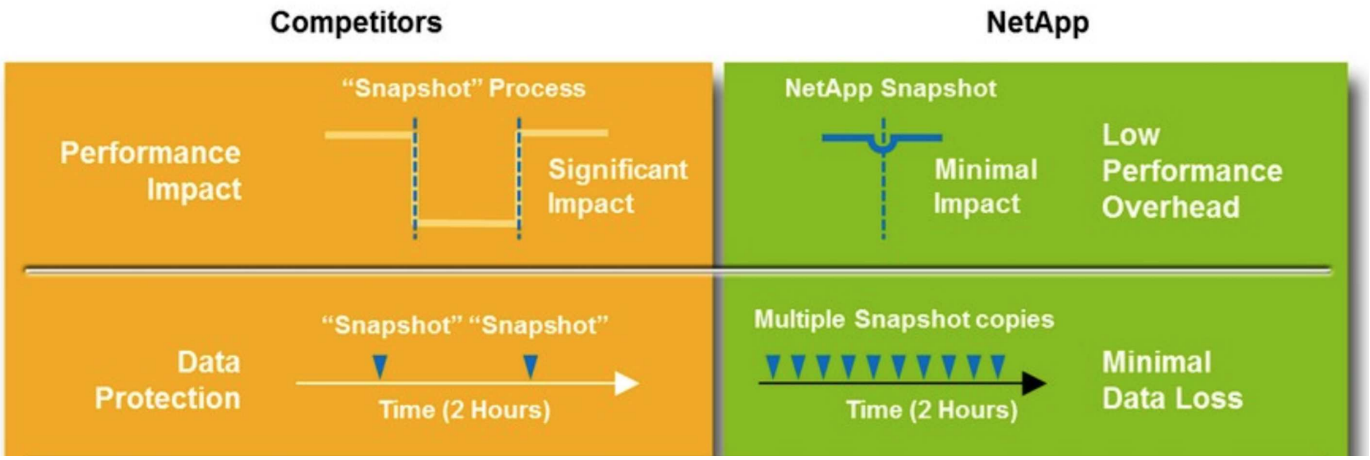


Figure 1a Figure 1b Figure 1c

A snapshot is taken in 1a. In 1b, changed data is written to a new block and the pointer is updated, but the snapshot pointer still points to the old block, giving you a live view of the data and an historical view. Another snapshot is taken in 1c and you now have access to 3 generations of your data without taking up the disk space that 3 unique copies would require; live, snapshot 2 and snapshot 1 in order of age.

A NetApp Cloud Volume Snapshot copy incurs no performance overhead; users can comfortably store up to 255 Snapshot copies per volume, all of which are accessible as read-only and online versions of the data. Read-only via the ".snapshot" directory located at the root of the filesystem and online from a cloned volume today and in-place restore in the near future. So what about user experience when working with the cloned volume? Because of the pointer system, use of the cloned volume should provide a comparable experience to the use of the parent volume itself.

NetApp Snapshot comparison



About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of data across cloud and on-premises environments to accelerate digital transformation. We empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize operations. For more information, visit: www.netapp.com #DataDriven

Oracle in the Clouds

NetApp Cloud Volumes Service For AWS

A review of Oracle performance as it pertains to the NetApp Cloud Volumes Service in AWS.

Authors: Ali Aiello, Joe Carter, Chad Morgenstern

If you are thinking about your data in the clouds, consider that the NetApp Cloud Volumes Service is about more than just NFS or SMB. With the NetApp Cloud Volumes Service, you can rest assured that your data is durable, encrypted, highly available and performant.

Please see the blog NetApp Cloud Volumes, Not Your Mothers File Service for details regarding how NetApp Cloud Volumes approaches durability, encryption, and availability as well as some really nice details regarding the Cloud Volumes snapshot technology.

The remainder of blog focuses on Oracle performance. Oracle is especially well suited for inclusion in the NetApp Cloud Volumes Service. Beyond the fact that your data is protected by 8 nines of durability and is accessible across all the availability zones within a given region, Oracle's Direct NFS client seems almost purpose built for Amazon's AWS architecture. More on that later.

Performance

When performing evaluations of database workloads, always keep in mind the impact of server memory. Ideally queries find their data resident therein as latency from memory is always going to be orders of magnitude lower than any "disk" query. What does this mean for you? While storage latency matters, memory hit percentage matters more. When database administrators say that they need a latency of X, keep in mind what they are talking about.

With that said…

The Oracle Direct NFS client spawns many network sessions to each NetApp Cloud volume, doing so as load

demands. The vast number of network sessions brings the potential for a significant amount of throughput for the database. Far greater than a single network session can provide in AWS.

To test Oracle's ability to generate load, NetApp ran a series of proofs of concept across various EC2 instance types, with various workload mixtures and volume counts. The SLOB2 workload generator drove the workload against a single Oracle 12.2c database. Keep in mind that at present AWS has no support for Oracle RAC, so single node testing was our only option.

The testing showed that the while any instance size is fine - run an instance size according to your need - Oracle and NetApp Cloud Volumes can consume the resources of even the largest instance. Towards that end, the graphics below focus on the c5.18xlarge instance type.

Workload : 100% Update

Amazon AWS imposes an architectural egress limit of 5Gbps for all I/O leaving the VPC. This limit defines the lower threshold in the graphic below, as seen in the 100% update workload represented by the green lines below. This workload amounts to 50% read and 50% write, the writes being the limiting factor. Though the two-volume test resulted in slightly lower application latency compared to the one-volume test, both volume counts converge at the same point the VPC egress limit which worked out to be just around 130,000 Oracle IOPS.

Workload : 100% Read

The 100% read workload represents the upper limit for the Oracle Database running against the NetApp Cloud Volumes Service in AWS. The pure read workload scenario, as shown by the red lines, is limited only by the instance type and the amount of network and CPU resources available to the instance. Just as the 100% update workload discussed above, though the two-volume configuration resulted in lower application latency, the tested configurations converged at the instance resource limit. In this case just shy of 300,000 Oracle IOPS.

Workload : 75% Read 25% Update

The third workload, shown in purple, is a 75% Oracle read and 25% Oracle update or read-modify-write workload. This scenario performed exceptionally well be it with one or two volumes, though the two-volume scenario outperformed the other. Many database administrators prefer the simplicity of a single volume database, 200,000 IOPS ought to satisfy all but the most demanding of database needs and doing so with the simple layout desired by these dba's. For those needing more, additional volumes will enable greater throughput as shown below.

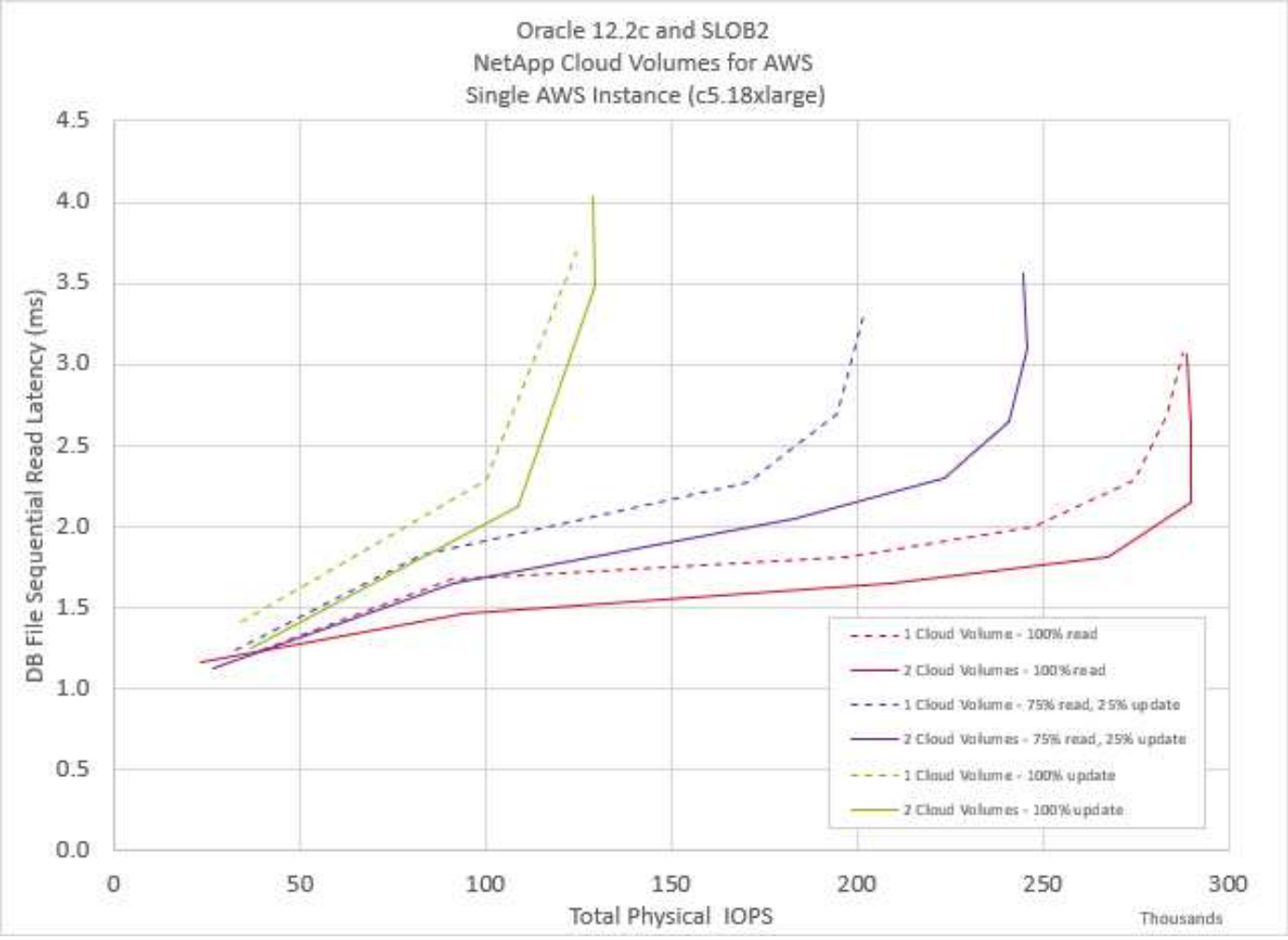


Figure 1. Oracle SLOB Test Graphic

Best Practices

- init.ora file
  - db_file_multiblock_read_countr
  - remove this option if present
- Redo block size:
  - Set to either 512 or 4KB, in general leave as default 512, unless recommended otherwise by App or Oracle.
  - If redo rates are greater than 50MBps, consider testing a 4KB block size
- Network considerations
  - Enable TCP timestamps, selective acknowledgement (SACK), and TCP window scaling on hosts
- Slot tables
  - sunrpc.tcp_slot_table_entries = 128
  - sunrpc.tcp_max_slot_table_entries = 65536
- Mount options

File Type	Mount Options

ADR_home	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536
Oracle Home	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr
Control Files	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr
Redo Logs	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr
Datafiles	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr

About NetApp

NetApp is the data authority for hybrid cloud.We provide a full range of hybrid cloud data services that simplify management of data across cloud and on-premises environments to accelerate digital transformation. We empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize operations. For more information, visit: www.netapp.com #DataDriven

MySQL in the Clouds

NetApp Cloud Volumes Service For AWS

A review of MySQL performance as it pertains to the NetApp Cloud Volumes Service in AWS.

Authors: Ali Aiello, Chad Morgenstern, Ron Pratt

Are looking for consistently good performance for your MySQL database - I mean, really, who is going to answer no to that question? With or without snapshots, whether you are accessing the primary database or a snapshotted copy, you can expect excellent, consistent performance from the NetApp Cloud Volumes Service. Performance is of course not the whole story. Databases need durability, data is a crown jewel of any enterprise. Consumers demand the protection against theft provided by encryption, we've got that too, its managed by the service. Add to these the advantage of accessing database volumes from any availability zone within the region – without the need to replicate to make this possible – and you'll find that the NetApp Cloud Volumes Service is the ideal solution for your MySQL needs.

Please see the blog NetApp Cloud Volumes, Not Your Mothers File Service for details regarding how NetApp Cloud Volumes approaches durability, encryption, and availability as well as some really nice details regarding the Cloud Volumes snapshot technology.

The remainder of this blog focuses on MySQL performance.

Performance

When performing evaluations of database workloads, always keep in mind the impact of server memory. Ideally queries find their data resident therein as latency from memory is always going to be orders of magnitude lower than any "disk" query. What does this mean for you? While storage latency matters, memory hit percentage matters more. When database administrators say that they need a latency of X, keep in mind what they are talking about.

With that said…

The Workload Generator

To test MySQL with cloud volumes, NetApp used the TPC Benchmark C workload generator. TPCC is an industry standard online transaction processing (OLTP) benchmark that leverages actual MySQL databases and their I/O paths. Workload generators that leverage real applications are always preferred over more synthetic generators such as Vdbench, Iozone, and heaven forbid dd, tar, cpio, or cp. TPCC standardized on an 80/20 read:write workload, the test results in this section are based there on.

The Scenarios

Two scenarios were tested in this environment, the first evaluated the capabilities of a single instance to drive MySQL I/O, while the second set out to determine the edges of a single cloud volume. The results of both scenarios are shown in the graphic below. The gray line represents the single-instance and blue the multi-instance environment. Please note that the latency reported in the graphics below represent storage latency as reported by the database.

The Results

Run against the single instance, TPCC generated approximately 4.5Gbps worth of I/O which approaches the Amazon Web Service inter-VPC limit imposed per network connection. Run against multiple instances, TPCC generated just about 16Gbps of throughput against a single Cloud Volume - which is pretty darned cool.

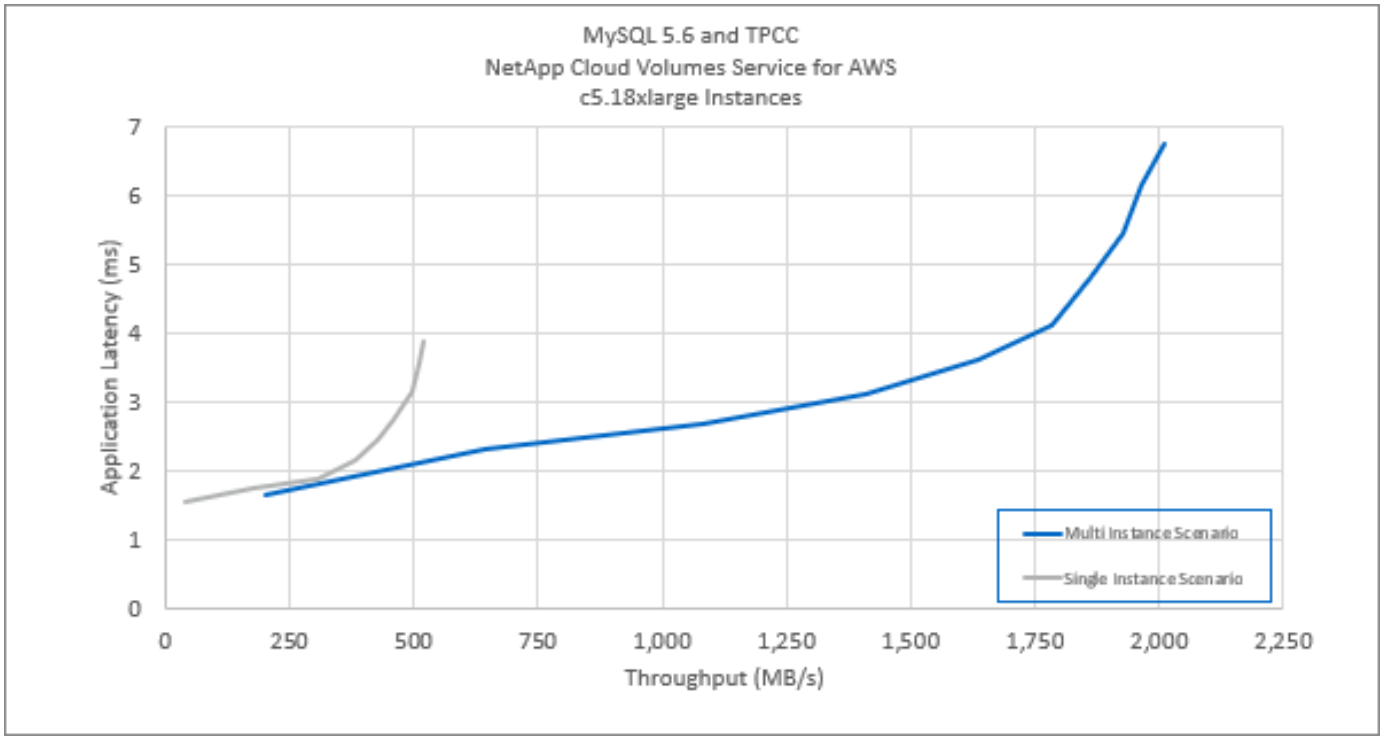


Figure 2. MySQL TPCC Test Graphic

Best Practices

The following parameters were placed in the MySQL /etc/my.cnf configuration file.

- [MySQLd]
  - innodb_buffer_pool_size=23622320128
  - innodb_log_buffer_size=4294967295
  - innodb_log_file_size=1073741824

- innodb_flush_log_at_trx_commit=2
- innodb_open_files=4096
- innodb_page_size=4096
- innodb_read_io_threads=64
- innodb_write_io_threads=64
- performance_schema
- innodb_doublewrite=0;
- max_connections=1000
- innodb_thread_concurrency=128
- innodb_max_dirty_pages_pct=0

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of data across cloud and on-premises environments to accelerate digital transformation. NetApp empowers global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize operations. For more information, visit https://www.netapp.com/#DataDriven

Cloud Volumes ONTAP

Blogs

NetApp Cloud Volumes ONTAP, A.K.A. ONTAP Cloud

Improved Performance Through Enhanced Networking Support

If you haven't heard the news, NetApp ONTAP Cloud has been rebranded as NetApp Cloud Volumes ONTAP. Along with the change comes exciting news of the release of support for ONTAP 9.4 – which in Amazon brings significant increases in front end performance. How cool is that, for once something really is both "New and Improved."

The most significant addition to the ONTAP 9.4 release for Cloud Volumes ONTAP is the inclusion of support for enhanced networking for the software defined storage system. As described by Amazon:

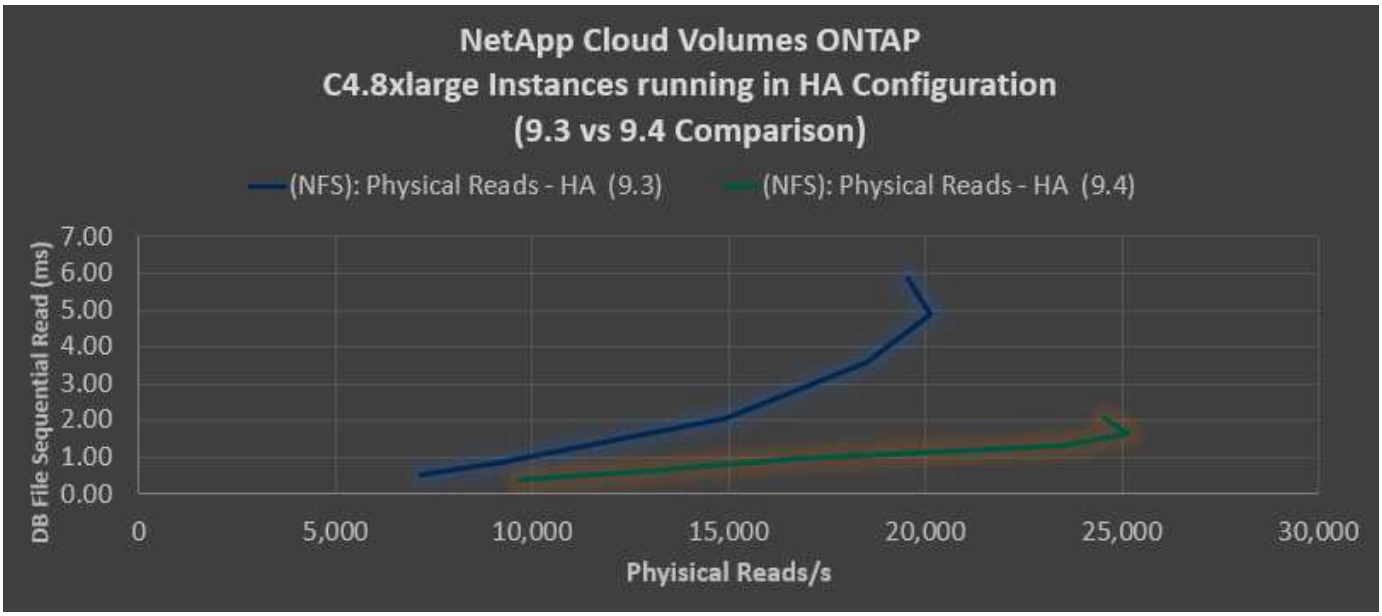
"Enhanced networking uses single root I/O virtualization (SR-IOV) to provide high-performance networking capabilities on supported instance types. SR-IOV is a method of device virtualization that provides higher I/O performance and lower CPU utilization when compared to traditional virtualized network interfaces. Enhanced networking provides higher bandwidth, higher packet per second (PPS) performance, and consistently lower inter-instance latencies. There is no additional charge for using enhanced networking."

Keep in mind that there are two types of enhanced networking interfaces supported by AWS and that NetApp supports at present the former (VF).

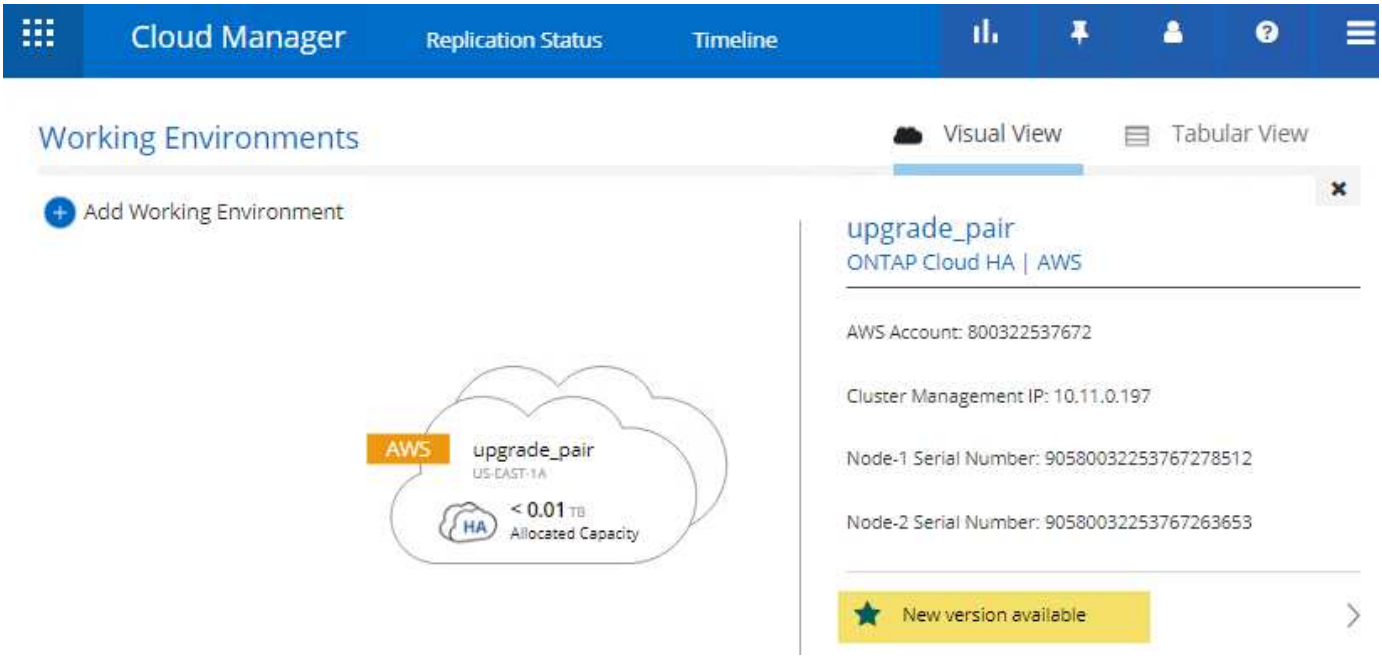
1. Intel 85299 Virtual Function (VF) interface which is supported within Xen hypervisor
2. Elastic network adapter supported from within KVM

The following graph calls out the increase in performance introduced by upgrading from ONTAP 9.3 to ONTAP 9.4 on an active-passive HA pair based on the c4.8xlarge instance type. Using the SLOB2 workload generator and Oracle 12.2c we performed a 75% read 25% update workload using an 8K database block size. While the

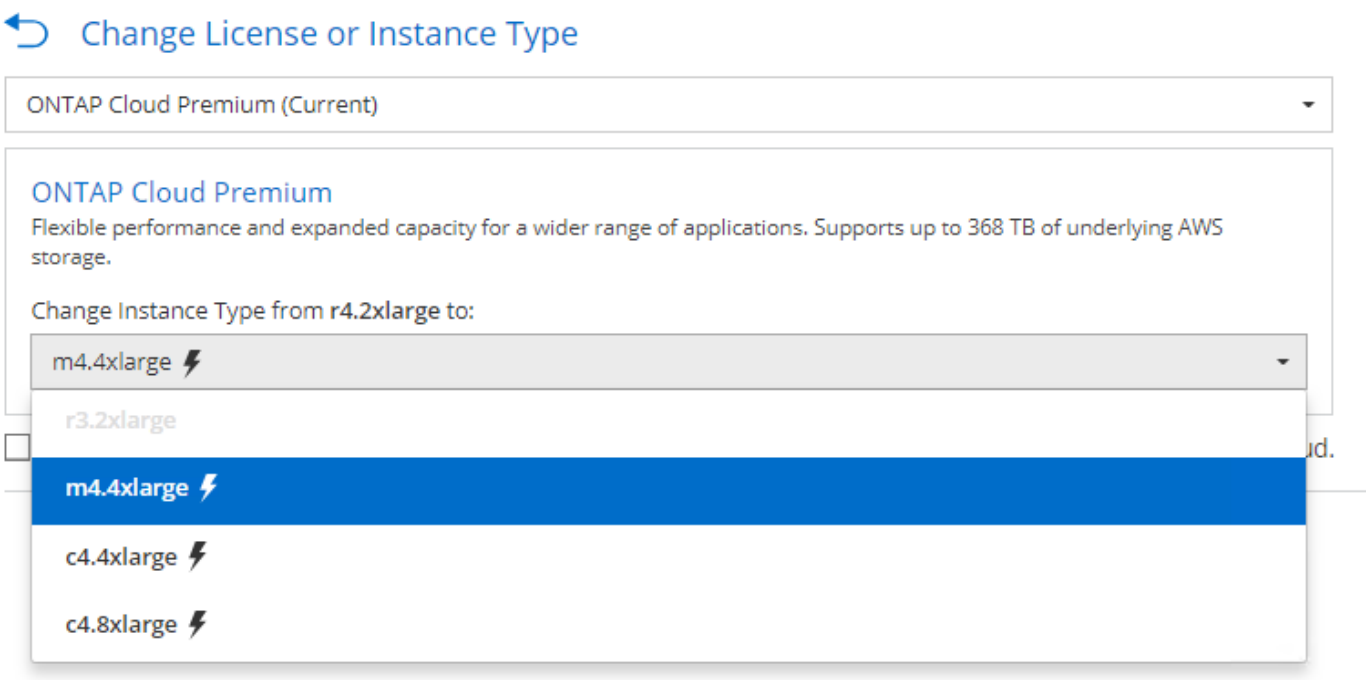
read I/O rate increased 25% simply by upgrading an Cloud Volumes ONTAP cluster, the latency reduction is even more impressive.



The upgrade path from ONTAP 9.3 to ONTAP 9.4 is a breeze with OnCommand Cloud Manager (OCCM). OCCM takes care of everything from upgrading the Cloud Volumes ONTAP nodes without downtime (assuming your environment is running in an HA configuration) to enabling sriov-net-support on the instances. Its as simple as clicking on the “New version available link” below and checking yes on a few “I have read and acknowledge the above.”



Not all instance types support enhanced networking, OCCM can help you there too by showing which instance types support this enchantment (those with the lightning bolt as shown below) as well as managing the process of instance type change in case you are interested in such. As the upgrade process, instance type change may be performed non-disruptively if running in an HA configuration.



About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of data across cloud and on-premises environments to accelerate digital transformation. We empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize operations. For more information, visit: www.netapp.com #DataDriven

NetApp Private Storage

Blogs

GDPR and Hybrid Cloud: A More Perfect Union

With the European Union's General Data Protection Regulation (GDPR) now in full swing, chances are that some area of your business must rapidly alter how it manages the personal information and data of EU persons. With potential fines of up to 4 percent of global revenue for a GDPR violation, it is behooved of every business to ensure compliance with these regulations. NetApp Hybrid Cloud can provide your company the competitive advantage of accelerating time to market in this highly regulated space.

By leveraging a hybrid cloud solution, organizations mitigate compliance risks associated with GDPR. Businesses may decide to leverage the cloud for less sensitive data all while storing more sensitive data on prem. The NetApp portfolio may be leveraged in several different ways to achieve the desired level of resiliency, cost-savings, and compute capacity while still maintaining compliance. Consider the following two scenarios:

1. **On-Prem All Flash FAS (AFF) and Cloud Volumes Service**
   A NetApp AFF A800 cluster delivers ultra-low latency of below 200 microsecond and massive throughput of up to 300 GB/s powered by a combination of NVMe SSDs and NVMe/FC connectivity. In addition to being able to run your apps with unprecedented speed and scale you can maintain full control of the data on-prem.
   NetApp Cloud Volumes offers customers a fully-managed, cloud-native file storage service that integrates with on-prem AFF system via NetApp CloudSync. This means you can offload less secure data into the cloud seamlessly. Furthermore, if you have archival regulation requiring long term retention, you may choose to move content to an even lower, long-term storage tier reducing cost and maintaining even the most stringent record-keeping requirements.
2. **NetApp Private Storage (NPS) for Cloud**
   Each NetApp Private Storage for Cloud solution consists of a NetApp AFF storage system. The NPS systems are collocated in an Equinix data center and are directly connected with a service provider's cloud compute using a dedicated, high-bandwidth, low-latency network connection. Leverage the scalability of cloud compute while honoring data privacy, regulatory, and sovereignty requirements. NPS for Cloud is based on the NetApp ONTAP® storage operating system, offering a common storage management framework, or data fabric, with your on-premises data center. Using a single known operating system, you secure your regulated data on-prem and your less stringent data in the cloud.

The flexibility built into the NetApp portfolio enables customers to meet GDPR requirements without compromising on cost, speed or resiliency. Deploy as your needs demand.

Industries

HealthCare

Genomics

Technical Reports

NetApp And PetaGene: Accelerating the Genomics Revolution

Authors: Karthik Chinnathambi, Mary Hayes, David LaBrosse, Chad Morgenstern, Justin Spears

The objective of this document is to explain how NetApp and PetaGene 'technology' helps to accelerate genomic data analysis and bio medical discoveries. It explores how our cloud and on-premise based solutions can simplify genomic research collaboration, reduce research IT cost, and shorten genomic analysis time.

Abstract

The goal is to achieve true "precision medicine" where doctors and other care givers can provide customized treatments and medications based upon an individual's unique genetic profile. While the 'old' healthcare treatment model of "one-size-fits-all" is still practiced, many clinicians believe that "personalized" care will become the standard around the world.

Benefits Summary

Below is a brief list of the benefits which the NetApp and PetaGene collaboration brings to genomic researchers and bio medical clinicians. We are proud to provide technology and tools which can accelerate discoveries and can improve patient care.

- **Smaller Files = Faster Genomic Data Flows** --- NetApp engineers recently confirmed that when PetaGene software shrinks the large genomic files into smaller, more portable files, there are immediate data flow benefits. For example, we verified that the smaller files can be transferred more rapidly across the NetApp Data Fabric. Faster genomic data benefits all stakeholders
- **5X Higher Data Efficiencies** --- Another benefit which NetApp engineers verified during the testing is higher levels of data efficiencies. After running the PetaSuite data compression software, NetApp resources reported as much as 5X improvement in space efficiency.
- **Increase in Research Collaboration** --- Having smaller files to manage makes it easier and faster to share genomic files between labs and research facilities across campus --- or even around the globe.
- **Cloud Enablement (AWS, Google, Azure, and more)** --- NetApp and PetaGene provide multiple options for managing genomic data in cloud environments like AWS, Google, and others. Genomic files can be seamlessly and securely moved to and from the cloud to support a variety of cloud-based workflows. Further, cold data can be tiered to object storage using FabricPool, freeing performance tiers for new sequencing projects.
- **Interoperability of Files without Decompression** --- Another benefit is that PetaSuite allows researchers to retain interoperability of existing workflows and formats. Specifically, PetaSuite software allows researchers and clinicians to continue using FASTQ and BAM file representations in their existing tools and pipelines, without needing to decompress first.
- **Reduce Operational Costs via Data Fabric - Life Cycle Management** --- PetaGene's software also helps to decrease research operations costs as it literally shrinks the amount of data which needs to be managed at different stages of its life cycle. Not only do the smaller genomic files enable more efficient data transfers, but they also accelerate data tiering. For example, some research data may need to be

archived for short-term (or long-term) access by genomic researchers. Decreasing the size of the files via PetaSuite, PetaGene's compression software, can save costs and time --- no matter if the data is in a cloud or on-premise. The NetApp Data Fabric model allows researchers to seamlessly migrate genomic data --- where it's needed --- when it's needed locally or globally.

Introduction

Scientists and bioinformaticians have long sought ways to reduce the size of the large genomic datasets by using a combination of data compression and reduction techniques. In the past, raw sequencer output had often been stored for extended periods, while bioinformaticians carried out the complex tasks needed in the assembly and alignment of the sequencing data. With these steps complete, the data could be used in variant calling and interpretation which are the vital steps in understanding gene expression and disease.

Today this process is highly automated and has been greatly accelerated through a combination of parallel processing and the availability of reference genomes. There are now a number of compressed genomic file formats that reduce the size of an individually stored genome down to a few tens of gigabytes; work that previously took months or years can now be turned around in little more than a day. This has greatly improved the ability of bioinformaticians to work with and transfer data to clinicians in an efficient and timely manner.

We are now entering the world of personalized or precision medicine. Faster sequencing and more compact datasets have increased the number of individual sequences that can be performed. Individual patient genomes and even their individual diseases, typically cancers, can be sequenced (tumors can and do have their own genetic makeup that can differ from that of the patient). This brings about great hope and opportunity for new insights, while at the same time increasing the pressure on data capacity.

Amplify Storage Efficiency

PetaGene's software addresses the challenges caused by growing volumes of genomics data. PetaSuite is a set of scalable complementary software tools that significantly reduce the size and cost of NGS data for storage and transfer. It achieves up to 5x reduction *(as per NetApp testing)* in both storage costs and data transfer times compared to BAM and gzipped FASTQ files. PetaSuite transparently integrates with existing storage infrastructure and bioinformatics pipelines, while PetaLink provides a powerful virtual file access system. PetaLink produces a high performance virtual file view of the compressed file. This virtual file can then be transparently used just like the original file by Linux toolchains, pipelines and genome browsers.

Observed Compression Ratios

Maximum Compression Ratio	Minimum Compression Ratio	Average Compression Ratio	Std. Deviation
4.50	**1.52**	**2.27**	**0.67**

Tested on 70 Genomic Data files

Increase Data Mobility

NetApp SnapMirror and CloudSync make it easy to move genomic data across various cloud environments and make it readily available where it is needed. This capability is further enhanced with PetaGene's unique data compression technique. Three reference designs can be leveraged to efficiently store and access genomic data, improving performance and reducing cost.

At a high level, the overall work flow is as such:

1. Raw genetic data is read by the sequencer, the results are written out in BCL format to an ONTAP cluster.

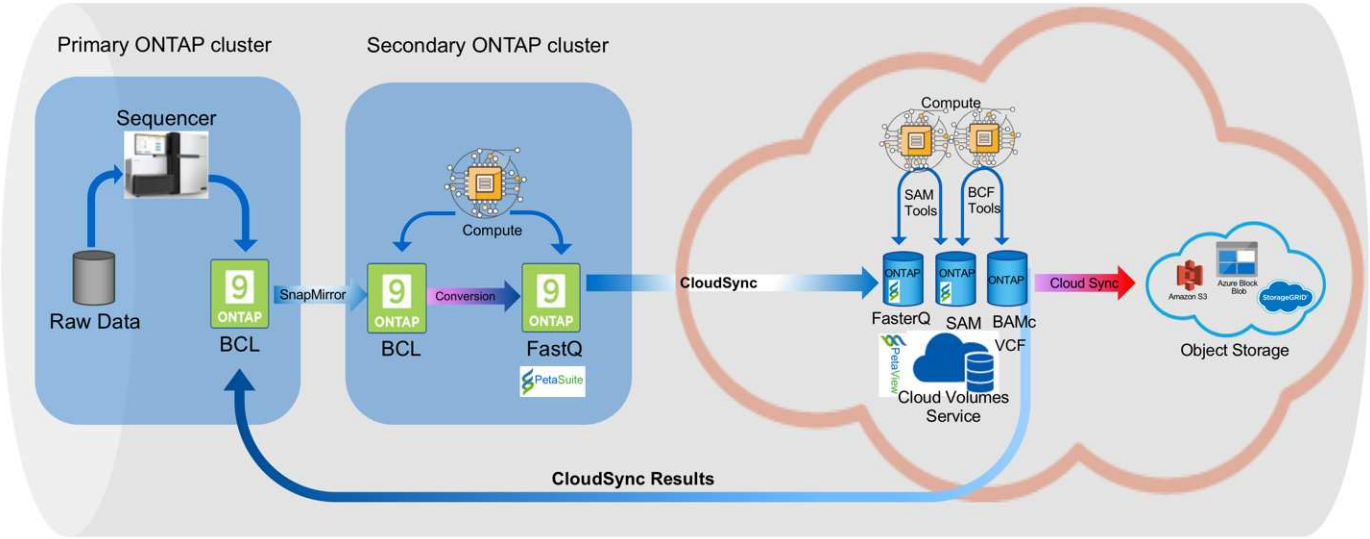
2. As many companies prefer to split conversion of BCL to FASTQ and BAM formats across separate environments, the BCL files are SnapMirrored into a second ONTAP cluster. Though not strictly necessary, this convention was followed in the reference design.
3. At the secondary site, the BCL files are processed and converted into FASTQ or BAM format.
4. With the BAM and FASTQ files in place, PetaSuite compresses the files converting to BAM to CRAM and FASTQ to FasterQ formats respectively. The space savings discussed above are realized at this stage.
5. The aforementioned compressed files are transferred from the secondary site to either volumes within the NetApp Cloud Volumes Service, a NetApp Cloud Volumes ONTAP Cluster or a NetApp Private Storage (NPS) Cluster. The transfer mechanism is dependent upon the target – CloudSync if NetApp Cloud Volumes Service or SnapMirror if NetApp Cloud Volumes ONTAP or NPS.
6. All post-processing activity is performed in one of the various hyperscaler environments and all file interactions occur against the NetApp storage target.
7. Resulting datasets may be replicated back to the origin data center as desired using either CloudSync or SnapMirror depending on the cloud solution.
8. Aged files may be archived to object storage either in the public cloud [AWS S3, Azure Blob, or a Storage Grid Webscale target] or on premises using StorageGRID Webscale. NetApp CloudSync or FabricPool are recommended to handle the archival process.

Speed, Scale and Simplicity with Cloud Volumes Service

NetApp Cloud Volumes Service is a cloud-native file storage service based on proven NetApp technology. This offering combines enterprise class storage with the simplicity and flexibility of the cloud, resulting in the ability to take your operation from 0TB to 100TB in less than 10 seconds. NetApp Cloud Volumes Service supports the NFS v3 and NFS v4 protocols along with SMB.

For simplified replication into and out of NetApp Cloud Volumes, CloudSync is an intuitive replication and synchronization service. This software-as-a-service (SaaS) offering enables customers to transfer and synchronize data between source and destination of any type or formats, in the cloud or on premises. CloudSync supports NAS data (NFS and SMB), EFS, Amazon S3 and NetApp StorageGRID Webscale Appliance.

Figure 1: Cloud Volumes Service reference design

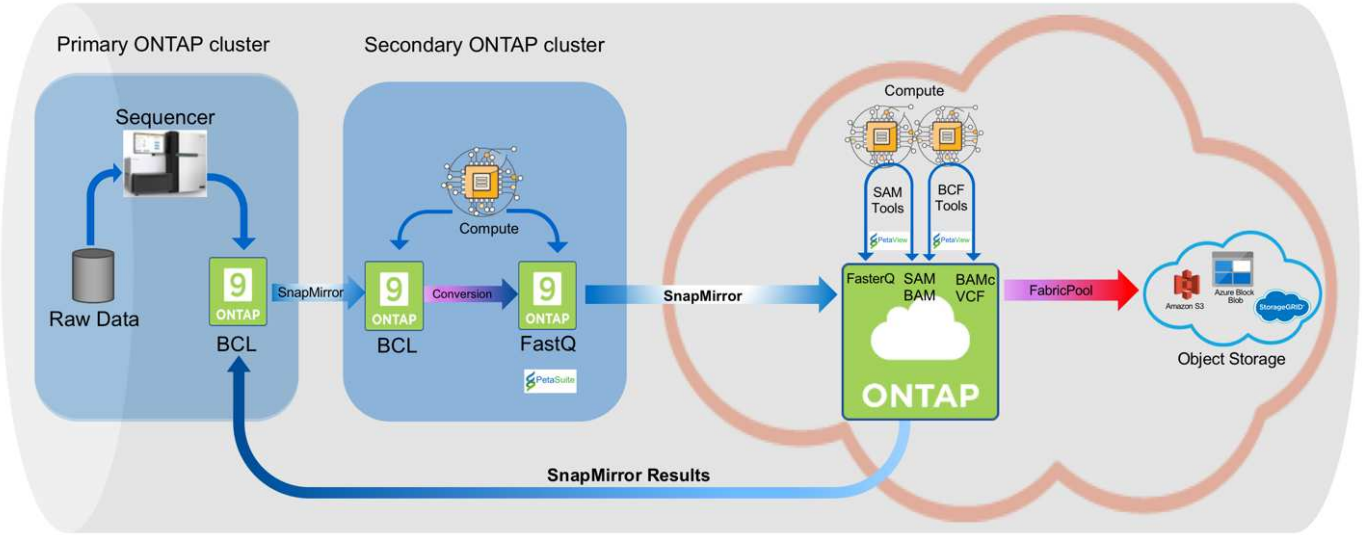


Control, Protection and Efficiency with Cloud Volumes ONTAP

NetApp Cloud Volumes ONTAP delivers enterprise control, protection, and efficiency to your data with the flexibility of the cloud. Cloud Volumes ONTAP, a software-defined data management service built on the NetApp ONTAP 9 storage operating system, provides a superior universal storage platform that addresses most cloud data needs. Having the same storage operating system in the cloud and on premises delivers the value of a Data Fabric without having to train IT staff in all-new methods to manage data. The SnapMirror features of ONTAP offer a bandwidth efficient data replication and transfer mechanism between clouds and to or from a datacenter.

Cloud Volumes ONTAP provides a data storage solution that fits many different customer requirements –from disaster recovery, development, and test environments to critical applications that require highly available non-disruptive operation, such as production business applications and file services using NFS, SMB, and iSCSI. Setup and management of the Cloud Volumes ONTAP environment is simple and intuitive with NetApp OnCommand Cloud Manager web interface.

Figure 2: Cloud Volumes ONTAP reference design

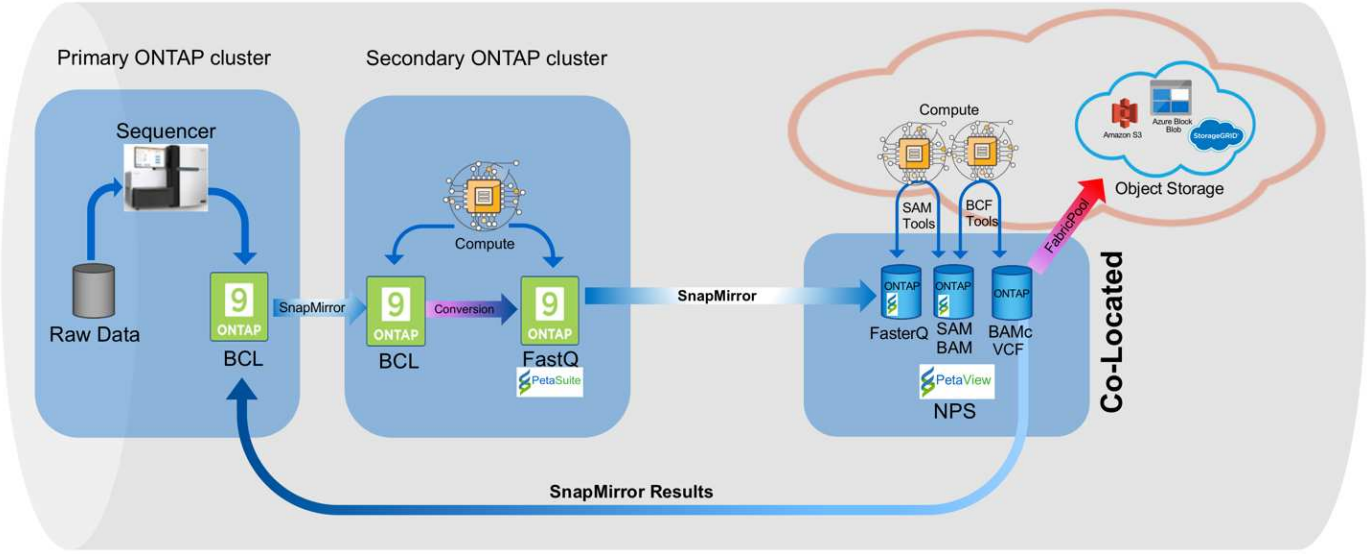


Freedom and Flexibility with NetApp Private Storage

NetApp Private Storage (NPS) is a cloud-connected storage solution that puts data near the cloud, providing the freedom and flexibility to run your application or workload on cloud compute while maintaining complete control of your data. NPS connectivity options allow you to choose from an expanding global network of cloud service providers, including Amazon Web Services, Google Cloud Platform, IBM Cloud and Microsoft Azure. Using NPS you can easily ensure compliance with HIPAA, GDPR or any other regulatory requirement.

With NPS, your NetApp storage is housed in co-located cloud-connected data centers, next to major networks and in close proximity to all major clouds. Establishing secure, dedicated, high-speed connections to all those clouds is quick and easy, with the added advantage of enhanced performance and reduced cost by bypassing the internet. NetApp makes it easy to move data between clouds and any NetApp data management infrastructure, including public, private, and hybrid clouds. SnapMirror technology provides support for applications to fail over to a secondary system and continue operating, as well as the capability to fail back to the primary location later.

Figure 3: NetApp Private Storage (NPS) design



Technology Comparison

	NetApp Private Storage	Cloud Volumes ONTAP	Cloud Volumes Service
Legal restrictions	X		
Data mobility	X	X	X
Multi cloud	X		
Cloudy deployment		X	X
Cost model	Opex & CapEx models available	OpEx	OpEx
Simplicity			X
Feature currency	X		
Storage Tiering	X	X	X

About PetaGene

PetaGene started from a team of Cambridge University PhDs who were working to devise a novel approach to the problem of storing the rapidly growing data associated Genomics. Through Project PetaGene, they created the means to compress huge amounts of genomic data without compromising data quality. PetaGene's technology goes beyond normal storage data reduction techniques, using insight into the structure of Genomic Data to achieve its industry leading data reduction. For more information visit www.petagene.com or email sales@petagene.com

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of data across cloud and on-premises environments to accelerate digital transformation. We empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize operations. For more information, visit: www.netapp.com #DataDriven

Business Applications

Oracle

Blogs

Oracle in the Clouds

NetApp Cloud Volumes Service For AWS

A review of Oracle performance as it pertains to the NetApp Cloud Volumes Service in AWS.

Authors: Ali Aiello, Joe Carter, Chad Morgenstern

If you are thinking about your data in the clouds, consider that the NetApp Cloud Volumes Service is about more than just NFS or SMB. With the NetApp Cloud Volumes Service, you can rest assured that your data is durable, encrypted, highly available and performant.

Please see the blog NetApp Cloud Volumes, Not Your Mothers File Service for details regarding how NetApp Cloud Volumes approaches durability, encryption, and availability as well as some really nice details regarding the Cloud Volumes snapshot technology.

The remainder of blog focuses on Oracle performance. Oracle is especially well suited for inclusion in the NetApp Cloud Volumes Service. Beyond the fact that your data is protected by 8 nines of durability and is accessible across all the availability zones within a given region, Oracle's Direct NFS client seems almost purpose built for Amazon's AWS architecture. More on that later.

Performance

When performing evaluations of database workloads, always keep in mind the impact of server memory. Ideally queries find their data resident therein as latency from memory is always going to be orders of magnitude lower than any "disk" query. What does this mean for you? While storage latency matters, memory hit percentage matters more. When database administrators say that they need a latency of X, keep in mind what they are talking about.

With that said…

The Oracle Direct NFS client spawns many network sessions to each NetApp Cloud volume, doing so as load demands. The vast number of network sessions brings the potential for a significant amount of throughput for the database. Far greater than a single network session can provide in AWS.

To test Oracle's ability to generate load, NetApp ran a series of proofs of concept across various EC2 instance types, with various workload mixtures and volume counts. The SLOB2 workload generator drove the workload against a single Oracle 12.2c database. Keep in mind that at present AWS has no support for Oracle RAC, so single node testing was our only option.

The testing showed that the while any instance size is fine - run an instance size according to your need - Oracle and NetApp Cloud Volumes can consume the resources of even the largest instance. Towards that end, the graphics below focus on the c5.18xlarge instance type.

Workload : 100% Update

Amazon AWS imposes an architectural egress limit of 5Gbps for all I/O leaving the VPC. This limit defines the lower threshold in the graphic below, as seen in the 100% update workload represented by the green lines

below. This workload amounts to 50% read and 50% write, the writes being the limiting factor. Though the two-volume test resulted in slightly lower application latency compared to the one-volume test, both volume counts converge at the same point the VPC egress limit which worked out to be just around 130,000 Oracle IOPS.

Workload : 100% Read

The 100% read workload represents the upper limit for the Oracle Database running against the NetApp Cloud Volumes Service in AWS. The pure read workload scenario, as shown by the red lines, is limited only by the instance type and the amount of network and CPU resources available to the instance. Just as the 100% update workload discussed above, though the two-volume configuration resulted in lower application latency, the tested configurations converged at the instance resource limit. In this case just shy of 300,000 Oracle IOPS.

Workload : 75% Read 25% Update

The third workload, shown in purple, is a 75% Oracle read and 25% Oracle update or read-modify-write workload. This scenario performed exceptionally well be it with one or two volumes, though the two-volume scenario outperformed the other. Many database administrators prefer the simplicity of a single volume database, 200,000 IOPS ought to satisfy all but the most demanding of database needs and doing so with the simple layout desired by these dba's. For those needing more, additional volumes will enable greater throughput as shown below.

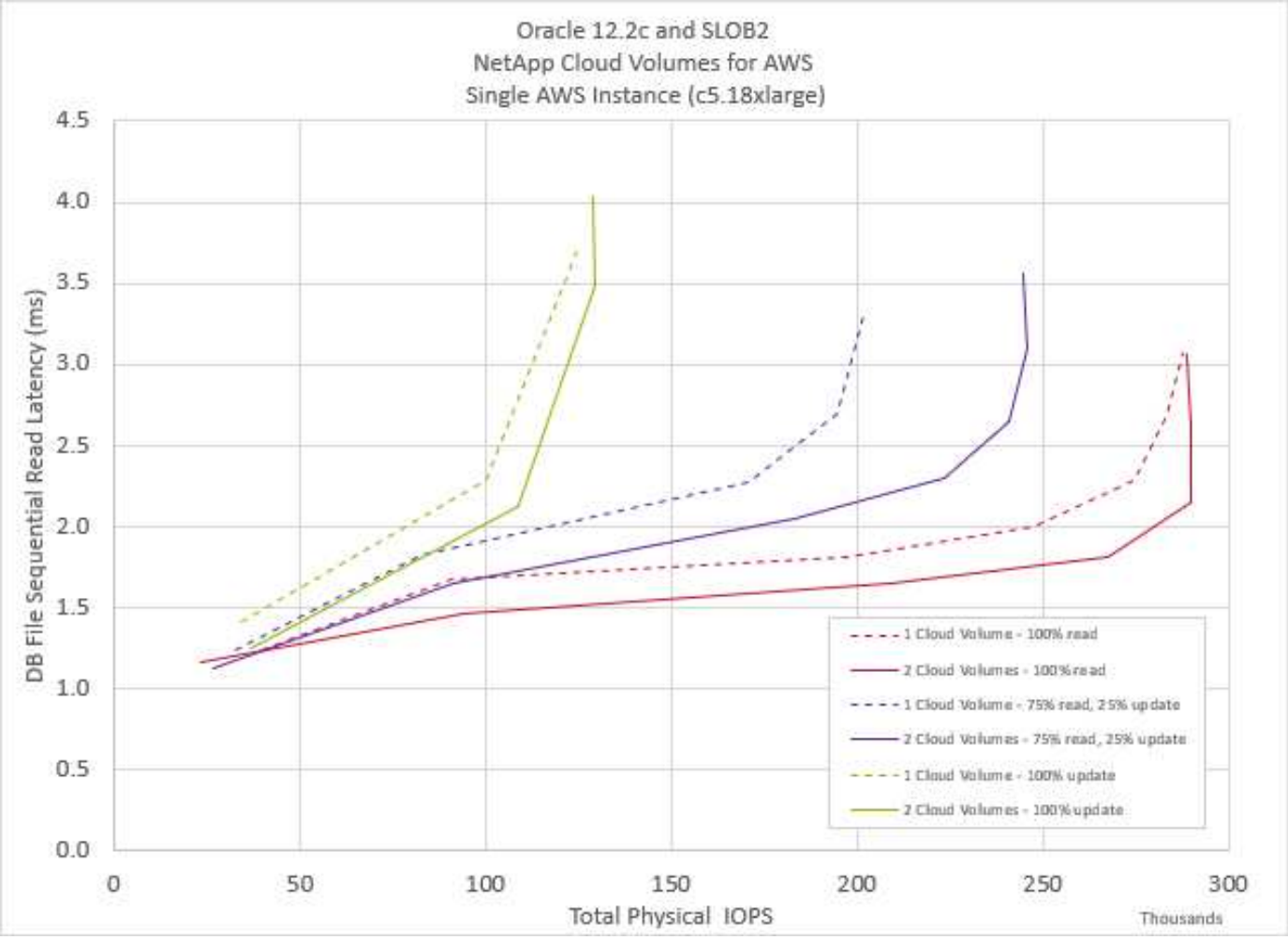


Figure 3. Oracle SLOB Test Graphic

Best Practices

- init.ora file
  - db_file_multiblock_read_countr
  - remove this option if present
- Redo block size:
  - Set to either 512 or 4KB, in general leave as default 512, unless recommended otherwise by App or Oracle.
  - If redo rates are greater than 50MBps, consider testing a 4KB block size
- Network considerations
  - Enable TCP timestamps, selective acknowledgement (SACK), and TCP window scaling on hosts
- Slot tables
  - sunrpc.tcp_slot_table_entries = 128
  - sunrpc.tcp_max_slot_table_entries = 65536
- Mount options

File Type	Mount Options
ADR_home	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536
Oracle Home	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr
Control Files	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr
Redo Logs	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr
Datafiles	rw,bg,hard,vers=3,proto=tcp,timeo=600,rsize=65536,wsize=65536,nointr

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of data across cloud and on-premises environments to accelerate digital transformation. NetApp empowers global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize operations. For more information, visit: www.netapp.com #DataDriven

Technical Reports

MySQL

Blogs

MySQL in the Clouds

NetApp Cloud Volumes Service For AWS

A review of MySQL performance as it pertains to the NetApp Cloud Volumes Service in AWS.

Authors: Ali Aiello, Chad Morgenstern, Ron Pratt

Are looking for consistently good performance for your MySQL database - I mean, really, who is going to answer no to that question? With or without snapshots, whether you are accessing the primary database or a snapshotted copy, you can expect excellent, consistent performance from the NetApp Cloud Volumes Service. Performance is of course not the whole story. Databases need durability, data is a crown jewel of any enterprise. Consumers demand the protection against theft provided by encryption, we've got that too, its managed by the service. Add to these the advantage of accessing database volumes from any availability zone within the region – without the need to replicate to make this possible – and you'll find that the NetApp Cloud Volumes Service is the ideal solution for your MySQL needs.

Please see the blog NetApp Cloud Volumes, Not Your Mothers File Service for details regarding how NetApp Cloud Volumes approaches durability, encryption, and availability as well as some really nice details regarding the Cloud Volumes snapshot technology.

The remainder of this blog focuses on MySQL performance.

Performance

When performing evaluations of database workloads, always keep in mind the impact of server memory. Ideally queries find their data resident therein as latency from memory is always going to be orders of magnitude lower than any "disk" query. What does this mean for you? While storage latency matters, memory hit percentage matters more. When database administrators say that they need a latency of X, keep in mind what they are talking about.

With that said…

The Workload Generator

To test MySQL with cloud volumes, NetApp used the TPC Benchmark C workload generator. TPCC is an industry standard online transaction processing (OLTP) benchmark that leverages actual MySQL databases and their I/O paths. Workload generators that leverage real applications are always preferred over more synthetic generators such as Vdbench, Iozone, and heaven forbid dd, tar, cpio, or cp. TPCC standardized on an 80/20 read:write workload, the test results in this section are based there on.

The Scenarios

Two scenarios were tested in this environment, the first evaluated the capabilities of a single instance to drive MySQL I/O, while the second set out to determine the edges of a single cloud volume. The results of both scenarios are shown in the graphic below. The gray line represents the single-instance and blue the multi-instance environment. Please note that the latency reported in the graphics below represent storage latency as reported by the database.

The Results

Run against the single instance, TPCC generated approximately 4.5Gbps worth of I/O which approaches the Amazon Web Service inter-VPC limit imposed per network connection. Run against multiple instances, TPCC generated just about 16Gbps of throughput against a single Cloud Volume - which is pretty darned cool.

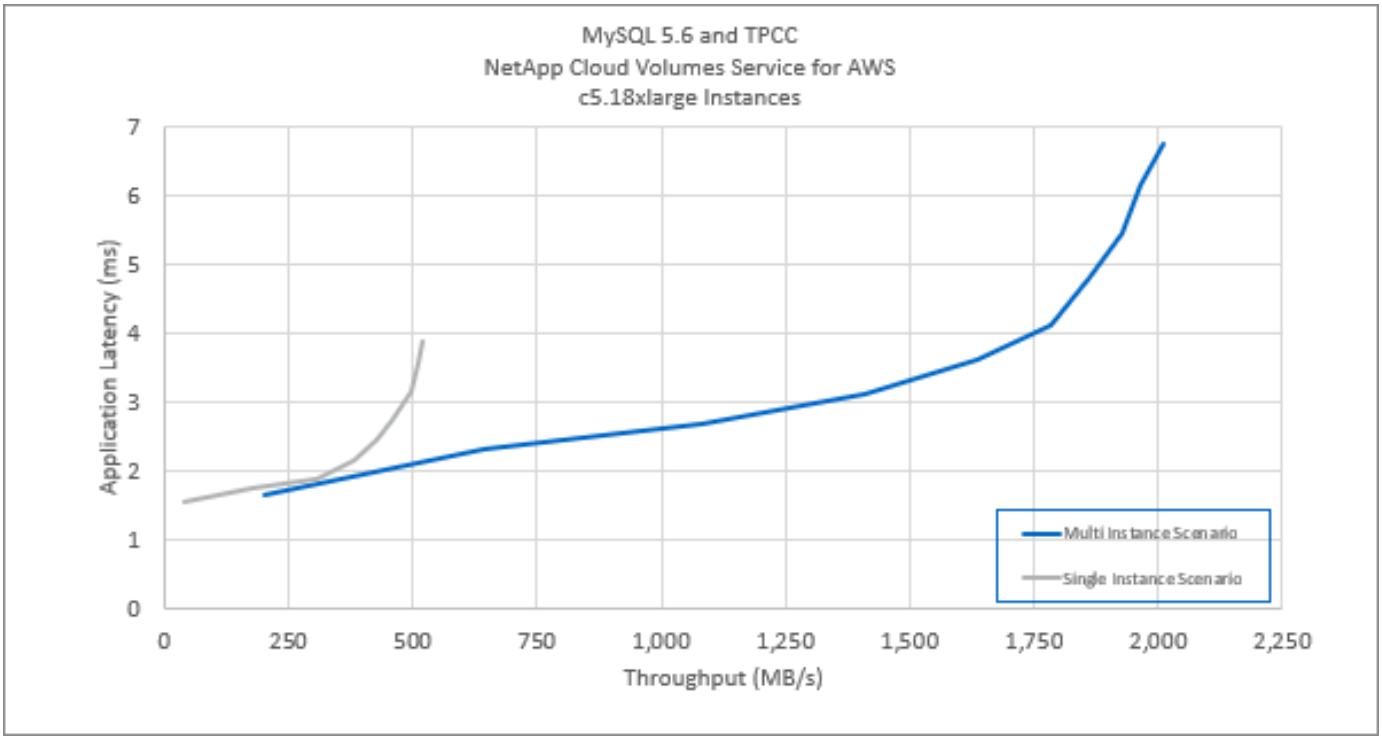


Figure 4. MySQL TPCC Test Graphic

Best Practices

The following parameters were placed in the MySQL /etc/my.cnf configuration file.

- [MySQLd]
  - innodb_buffer_pool_size=23622320128
  - innodb_log_buffer_size=4294967295
  - innodb_log_file_size=1073741824
  - innodb_flush_log_at_trx_commit=2
  - innodb_open_files=4096
  - innodb_page_size=4096
  - innodb_read_io_threads=64
  - innodb_write_io_threads=64
  - performance_schema
  - innodb_doublewrite=0;
  - max_connections=1000
  - innodb_thread_concurrency=128
  - innodb_max_dirty_pages_pct=0

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of data across cloud and on-premises environments to accelerate digital transformation. NetApp empowers global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize operations. For more information, visit: www.netapp.com #DataDriven

Copyright information

Trademark information